

Stat Comp Hw #5

Jason Rights

November 8, 2015

Problem 1:

```
setwd("C:/Users/Jason/Documents")
haart <- read.csv("haart.csv")
head(haart)
```

```
##   male age aids cd4baseline logvl  weight hemoglobin  init.reg init.date
## 1    1  25   0         NA     NA      NA          NA 3TC,AZT,EFV   7/1/03
## 2    1  49   0        143     NA  58.0608         11 3TC,AZT,EFV  11/23/04
## 3    1  42   1        102     NA  48.0816          1 3TC,AZT,EFV   4/30/03
## 4    0  33   0        107     NA  46.0000         NA 3TC,AZT,NVP   3/25/06
## 5    1  27   0         52     4      NA          NA 3TC,D4T,EFV   9/1/04
## 6    0  34   0        157     NA  54.8856         NA 3TC,AZT,NVP  12/2/03
##   last.visit death date.death
## 1    2/26/07    0      <NA>
## 2    2/22/08    0      <NA>
## 3   11/21/05    1    1/11/06
## 4     5/5/06    1     5/7/06
## 5   11/13/07    0      <NA>
## 6    2/28/08    0      <NA>
```

```
haart[,9] <- as.Date(haart[,9],format="%m/%d/%y")
haart[,10] <- as.Date(haart[,10],format="%m/%d/%y")
haart[,12] <- as.Date(haart[,12],format="%m/%d/%y")
```

```
haart[,13] <- format(haart[,9],"%Y")
table(haart[,13])
```

```
##
## 1998 2000 2001 2002 2003 2004 2005 2006 2007
##      1    5   17   60  270  292  207  104   44
```

```
haart[,14] <- as.numeric(haart[,12]-haart[,9])
haart[,15] <- 0
dim(haart[which(haart[,14]<365),])
```

```
## [1] 92 15
```

```
for(i in seq(nrow(haart))){
  if (is.na(haart[i,14])==FALSE && haart[i,14] < 365) haart[i,15] <- 1
}
sum(haart[,15])
```

```
## [1] 92
```

92 people died within one year.

```
##part 3
for(i in seq(nrow(haart))){
  haart[i,16] <- min(as.numeric(haart[i,10] - haart[i,9]), as.numeric(haart[i,12] - haart[i,9]),na.rm=T)
}

haart[,16] <- as.numeric(haart[,16])

head(haart)
```

```
##   male age aids cd4baseline logvl  weight hemoglobin  init.reg
## 1    1  25   0         NA      NA      NA          NA 3TC,AZT,EFV
## 2    1  49   0        143      NA  58.0608         11 3TC,AZT,EFV
## 3    1  42   1        102      NA  48.0816          1 3TC,AZT,EFV
## 4    0  33   0        107      NA  46.0000         NA 3TC,AZT,NVP
## 5    1  27   0         52       4      NA          NA 3TC,D4T,EFV
## 6    0  34   0        157      NA  54.8856         NA 3TC,AZT,NVP
##   init.date last.visit death date.death  V13 V14 V15 V16
## 1 2003-07-01 2007-02-26     0      <NA> 2003  NA   0 1336
## 2 2004-11-23 2008-02-22     0      <NA> 2004  NA   0 1186
## 3 2003-04-30 2005-11-21     1 2006-01-11 2003 987   0  936
## 4 2006-03-25 2006-05-05     1 2006-05-07 2006  43   1   41
## 5 2004-09-01 2007-11-13     0      <NA> 2004  NA   0 1168
## 6 2003-12-02 2008-02-28     0      <NA> 2003  NA   0 1549
```

```
for(i in seq(nrow(haart))){
  if (is.na(haart[i,16])==FALSE && haart[i,16] > 365) haart[i,16] <- 365
}

quantile(haart[,16],na.rm=T)
```

```
##      0%      25%      50%      75%     100%
##    0.00 320.75 365.00 365.00 365.00
```

```
##part 4
for(i in seq(nrow(haart))){
  haart[i,17] <- 0
  if (is.na(haart[i,12])==TRUE && haart[i,10]-haart[i,9] < 365) haart[i,17] <- 1
}
sum(haart[,17])
```

```
## [1] 173
```

173 records lost to followup.

```
##part 5

reg_list<-strsplit(as.character(haart[, 'init.reg']),',')
str(haart)
```

```
## 'data.frame':    1000 obs. of  17 variables:
## $ male          : int  1 1 1 0 1 0 0 1 1 1 ...
## $ age           : num  25 49 42 33 27 34 39 31 52 23 ...
## $ aids          : int  0 0 1 0 0 0 0 0 0 1 ...
## $ cd4baseline: int  NA 143 102 107 52 157 65 NA NA 3 ...
## $ logvl         : num  NA NA NA NA 4 ...
## $ weight        : num  NA 58.1 48.1 46 NA ...
## $ hemoglobin    : num  NA 11 1 NA NA NA 11 NA NA NA ...
## $ init.reg      : Factor w/ 47 levels "3TC,ABC,AZT",...: 10 10 10 17 19 17 17 10 1 27 ...
## $ init.date     : Date, format: "2003-07-01" "2004-11-23" ...
## $ last.visit    : Date, format: "2007-02-26" "2008-02-22" ...
## $ death         : int  0 0 1 1 0 0 0 0 0 1 ...
## $ date.death    : Date, format: NA NA ...
## $ V13           : chr  "2003" "2004" "2003" "2006" ...
## $ V14           : num  NA NA 987 43 NA NA NA NA NA 87 ...
## $ V15           : num  0 0 0 1 0 0 0 0 0 1 ...
## $ V16           : num  365 365 365 41 365 365 365 365 365 84 ...
## $ V17           : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
head(reg_list)
```

```
## [[1]]
## [1] "3TC" "AZT" "EFV"
##
## [[2]]
## [1] "3TC" "AZT" "EFV"
##
## [[3]]
## [1] "3TC" "AZT" "EFV"
##
## [[4]]
## [1] "3TC" "AZT" "NVP"
##
## [[5]]
## [1] "3TC" "D4T" "EFV"
##
## [[6]]
## [1] "3TC" "AZT" "NVP"
```

```
head(sapply(reg_list,function(x) 'D4T' %in% x))
```

```
## [1] FALSE FALSE FALSE FALSE TRUE FALSE
```

```
all_drugs <- unique(unlist(reg_list))
reg_drugs <- matrix(nrow=nrow(haart),ncol=length(all_drugs))
for(i in seq_along(all_drugs)) {
  # + makes this 1/0 instad of T/F
  reg_drugs[,i] <- +sapply(reg_list,function(x) all_drugs[i] %in% x)
}
colnames(reg_drugs) <- all_drugs
haart <- cbind(haart, reg_drugs)
```

```
reg_drugs <- as.data.frame(reg_drugs)
sapply(reg_drugs, sum)
```

```
## 3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20 ATV FPV
## 973 794 516 358 146 56 38 27 31 79 29 8 10 1 8 1 2 2
```

3TC, AZT, EFV, NVP, and D4T all occur over 100 times.

```
setwd("C:/Users/Jason/Documents")
haart <- read.csv("haart.csv")

haart2 <- read.csv("haart2.csv")
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'haart2.csv'
```

```
haart <- rbind(haart, haart2)
haart[,9] <- as.Date(haart[,9],format="%m/%d/%y")
haart[,10] <- as.Date(haart[,10],format="%m/%d/%y")
haart[,12] <- as.Date(haart[,12],format="%m/%d/%y")
haart[,13] <- format(haart[,9],"%Y")
table(haart[,13])
```

```
##
## 1998 2000 2001 2002 2003 2004 2005 2006 2007
## 1 5 17 61 272 292 207 104 45
```

```
haart[,14] <- as.numeric(haart[,12]-haart[,9])
haart[,15] <- 0
dim(haart[which(haart[,14]<365),])
```

```
## [1] 92 15
```

```
for(i in seq(nrow(haart))){
  if (is.na(haart[i,14])==FALSE && haart[i,14] < 365) haart[i,15] <- 1
}

for(i in seq(nrow(haart))){
  haart[i,16] <- min(as.numeric(haart[i,10] - haart[i,9]), as.numeric(haart[i,12] - haart[i,9]),na.rm=T)
}

haart[,16] <- as.numeric(haart[,16])

for(i in seq(nrow(haart))){
  if (is.na(haart[i,16])==FALSE && haart[i,16] > 365) haart[i,16] <- 365
}
```

```
##part 4
for(i in seq(nrow(haart))){
  haart[i,17] <- 0
  if (is.na(haart[i,12])==TRUE && haart[i,10]-haart[i,9] < 365) haart[i,17] <- 1
}
```

```
##part 5
```

```
reg_list<-strsplit(as.character(haart[, 'init.reg']), ',')
str(haart)
```

```
## 'data.frame': 1004 obs. of 17 variables:
## $ male : int 1 1 1 0 1 0 0 1 1 1 ...
## $ age : num 25 49 42 33 27 34 39 31 52 23 ...
## $ aids : int 0 0 1 0 0 0 0 0 0 1 ...
## $ cd4baseline: int NA 143 102 107 52 157 65 NA NA 3 ...
## $ logvl : num NA NA NA NA 4 ...
## $ weight : num NA 58.1 48.1 46 NA ...
## $ hemoglobin : num NA 11 1 NA NA NA 11 NA NA NA ...
## $ init.reg : Factor w/ 47 levels "3TC,ABC,AZT",...: 10 10 10 17 19 17 17 10 1 27 ...
## $ init.date : Date, format: "2003-07-01" "2004-11-23" ...
## $ last.visit : Date, format: "2007-02-26" "2008-02-22" ...
## $ death : int 0 0 1 1 0 0 0 0 0 1 ...
## $ date.death : Date, format: NA NA ...
## $ V13 : chr "2003" "2004" "2003" "2006" ...
## $ V14 : num NA NA 987 43 NA NA NA NA NA 87 ...
## $ V15 : num 0 0 0 1 0 0 0 0 0 1 ...
## $ V16 : num 365 365 365 41 365 365 365 365 84 ...
## $ V17 : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
head(reg_list)
```

```
## [[1]]
## [1] "3TC" "AZT" "EFV"
##
## [[2]]
## [1] "3TC" "AZT" "EFV"
##
## [[3]]
## [1] "3TC" "AZT" "EFV"
##
## [[4]]
## [1] "3TC" "AZT" "NVP"
##
## [[5]]
## [1] "3TC" "D4T" "EFV"
##
## [[6]]
## [1] "3TC" "AZT" "NVP"
```

```
head(sapply(reg_list,function(x) 'D4T' %in% x))
```

```
## [1] FALSE FALSE FALSE FALSE TRUE FALSE
```

```
all_drugs <- unique(unlist(reg_list))
reg_drugs <- matrix(nrow=nrow(haart),ncol=length(all_drugs))
for(i in seq_along(all_drugs)) {
  # + makes this 1/0 instad of T/F
  reg_drugs[,i] <- +sapply(reg_list,function(x) all_drugs[i] %in% x)
}
colnames(reg_drugs) <- all_drugs
haart <- cbind(haart, reg_drugs)
haart[1:5,]
```

```
##   male age aids cd4baseline logvl  weight hemoglobin  init.reg
## 1    1  25   0         NA    NA      NA          NA 3TC,AZT,EFV
## 2    1  49   0        143    NA  58.0608         11 3TC,AZT,EFV
## 3    1  42   1        102    NA  48.0816          1 3TC,AZT,EFV
## 4    0  33   0        107    NA  46.0000         NA 3TC,AZT,NVP
## 5    1  27   0         52     4     NA          NA 3TC,D4T,EFV
##   init.date last.visit death date.death  V13 V14 V15 V16 V17 3TC AZT EFV
## 1 2003-07-01 2007-02-26     0      <NA> 2003  NA   0 365   0   1   1   1
## 2 2004-11-23 2008-02-22     0      <NA> 2004  NA   0 365   0   1   1   1
## 3 2003-04-30 2005-11-21     1 2006-01-11 2003 987   0 365   0   1   1   1
## 4 2006-03-25 2006-05-05     1 2006-05-07 2006  43   1  41   0   1   1   0
## 5 2004-09-01 2007-11-13     0      <NA> 2004  NA   0 365   0   1   0   1
##   NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20 ATV FPV
## 1    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3    0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4    1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 5    0   1   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
haart[(nrow(haart)-4):nrow(haart),]
```

```
##      male      age aids cd4baseline  logvl  weight hemoglobin
## 1000    0 40.00000    1      131     NA  46.2672          8
## 1001    0 27.00000    0      232     NA     NA          NA
## 1002    1 38.72142    0      170     NA  84.0000          NA
## 1003    1 23.00000   NA      154  3.995635  65.5000         14
## 1004    0 31.00000    0      236     NA  45.8136          NA
##      init.reg  init.date last.visit death date.death  V13 V14 V15 V16
## 1000 3TC,D4T,NVP 2003-07-03 2008-02-29     0      <NA> 2003  NA   0 365
## 1001 3TC,AZT,NVP 2003-12-01 2004-01-05     0      <NA> 2003  NA   0 35
## 1002 3TC,AZT,NVP 2002-09-26 2004-03-29     0      <NA> 2002  NA   0 365
## 1003 3TC,DDI,EFV 2007-01-31 2007-04-16     0      <NA> 2007  NA   0 75
## 1004 3TC,D4T,NVP 2003-12-03 2007-10-11     0      <NA> 2003  NA   0 365
##      V17 3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20
## 1000    0   1   0   0   1   1   0   0   0   0   0   0   0   0   0   0
## 1001    1   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0
## 1002    0   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0
## 1003    1   1   0   1   0   0   0   1   0   0   0   0   0   0   0   0
```

```
## 1004  0  1  0  0  1  1  0  0  0  0  0  0  0  0  0  0  0
##      ATV FPV
## 1000  0  0
## 1001  0  0
## 1002  0  0
## 1003  0  0
## 1004  0  0
```

Problem 2:

```
setwd("C:/Users/Jason/Documents")
haart <- read.csv("haart.csv")
# Logistic function
logistic <- function(x) 1 / (1 + exp(-x))

haart.complete <- haart[c(4,6:7,11)]
haart.complete <- na.omit(haart.complete)

x <- haart.complete[c(1,2:3)]
y <- haart.complete[4]

dim(y)
```

```
## [1] 328  1
```

```
estimate_logistic <- function(x, y, MAX_ITER=10) {

  n <- dim(x)[1]
  k <- dim(x)[2]

  x <- as.matrix(cbind(rep(1, n), x))
  y <- as.matrix(y)

  # Initialize fitting parameters
  theta <- rep(0, k+1)

  J <- rep(0, MAX_ITER)

  for (i in 1:MAX_ITER) {

    # Calculate linear predictor
    z <- x %*% theta
    # Apply logit function
    h <- logistic(z)

    # Calculate gradient
    grad <- t((1/n)*x) %*% as.matrix(h - y)
    # Calculate Hessian
    H <- t((1/n)*x) %*% diag(array(h)) %*% diag(array(1-h)) %*% x

    # Calculate log likelihood
    J[i] <- (1/n) %*% sum(-y * log(h) - (1-y) * log(1-h))
```

```

# Newton's method
theta <- theta - solve(H) %*% grad
}

return(theta)
}

```

```
estimate_logistic(x, y)
```

```

##                [,1]
## rep(1, n)      3.576411744
## cd4baseline    0.002092582
## weight         -0.046210552
## hemoglobin     -0.350642786

```

Problem 3:

```

setwd("C:/Users/Jason/Documents")
addr <- read.delim("addr.txt",sep="\n",header=FALSE,stringsAsFactors=FALSE)
head(addr)

```

```

##
## 1      Bania   Thomas M.      725 Commonwealth Ave.   Boston  MA      02215
## 2 Barnaby      David   373 W. Geneva St.      Wms. Bay  WI      53191
## 3      Bausch   Judy    373 W. Geneva St.      Wms. Bay  WI      53191
## 4 Bolatto      Alberto 725 Commonwealth Ave.   Boston  MA      02215
## 5 Carlstrom    John   933 E. 56th St.      Chicago  IL      60637
## 6      Chamberlin Richard A. 111 Nowelo St.   Hilo    HI      96720

```

```

clean <- c(0,0,0,0,0,0)
for(i in 1:nrow(addr)){
  split <- unlist(strsplit(addr[i,],split=" "))
  rem <- split[nchar(split)>0]
  clean <- rbind(clean,rem)
}
clean <- clean[2:43,]
remstreet <- clean[,3]

street <- data.frame(character(0),character(0))
for(i in 1:nrow(addr)){
  split <- unlist(strsplit(remstreet[i],split=" "))
  rem <- split[nchar(split)>0]
  streetnum <- rem[1]
  streetnam <- rem[2:length(rem)]
  streetnam <- paste(streetnam,collapse = " ")
  street_ <- cbind(streetnum,streetnam)
  street <- rbind(street,street_)
}
address <- cbind(clean[,1:2],street,clean[,4:6])

```

```

## Warning in data.row.names(row.names, row si, i): some row.names duplicated:
## 2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,
## --> row.names NOT used

```



```
colnames(address) <- c("lastname", "firstname", "streetno", "streetname", "city", "state", "zip")
rownames(address) <- NULL
address
```

##	lastname	firstname	streetno			
## 1	Bania	Thomas M.	725 Commonwealth Ave.			
## 2	Barnaby	David	373 W. Geneva St.			
## 3	Bausch	Judy	373 W. Geneva St.			
## 4	Bolatto	Alberto	725 Commonwealth Ave.			
## 5	Carlstrom	John	933 E. 56th St.			
## 6	Chamberlin	Richard A.	111 Nowelo St.			
## 7	Chuss	Dave	2145 Sheridan Rd			
## 8	Davis	E. J.	933 E. 56th St.			
## 9	Depoy	Darren	174 W. 18th Ave.			
## 10	Griffin	Greg	5000 Forbes Ave.			
## 11	Halvorsen	Nils	933 E. 56th St.			
## 12	Harper	Al	373 W. Geneva St.			
## 13	Huang	Maohai	725 W. Commonwealth Ave.			
## 14	Ingalls	James G.	725 W. Commonwealth Ave.			
## 15	Jackson	James M.	725 W. Commonwealth Ave.			
## 16	Knudsen	Scott	373 W. Geneva St.			
## 17	Kovac	John	5640 S. Ellis Ave.			
## 18	Landsberg	Randy	5640 S. Ellis Ave.			
## 19	Lo	Kwok-Yung	1002 W. Green St.			
## 20	Loewenstein	Robert F.	373 W. Geneva St.			
## 21	Lynch	John	4201 Wilson Blvd			
## 22	Martini	Paul	174 W. 18th Ave.			
## 23	Meyer	Stephan	933 E. 56th St.			
## 24	Mrozek	Fred	373 W. Geneva St.			
## 25	Newcomb	Matt	5000 Forbes Ave.			
## 26	Novak	Giles	2145 Sheridan Rd			
## 27	Odalen	Nancy	373 W. Geneva St.			
## 28	Pernic	Dave	373 W. Geneva St.			
## 29	Pernic	Bob	373 W. Geneva St.			
## 30	Peterson	Jeffrey	5000 Forbes Ave.			
## 31	Pryke	Clem	933 E. 56th St.			
## 32	Rebull	Luisa	5640 S. Ellis Ave.			
## 33	Renbarger	Thomas	2145 Sheridan Rd			
## 34	Rottman	Joe	8730 W. Mountain View Ln			
## 35	Schartman	Ethan	933 E. 56th St.			
## 36	Spotz	Bob	373 W. Geneva St.			
## 37	Thoma	Mark	373 W. Geneva St.			
## 38	Walker	Chris	933 N. Cherry St.			
## 39	Wehrer	Cheryl	5000 Forbes Ave.			
## 40	Wirth	Jesse	373 W. Geneva St.			
## 41	Wright	Greg	791 Holmdel-Keyport Rd.			
## 42	Zingale	Michael	5640 S. Ellis Ave.			
##		streetname	city	state	zip	
## 1	NA	725 Commonwealth Ave.	Boston	MA	02215	
## 2	NA	373 W. Geneva St.	Wms. Bay	WI	53191	
## 3	NA	373 W. Geneva St.	Wms. Bay	WI	53191	
## 4	NA	725 Commonwealth Ave.	Boston	MA	02215	
## 5	NA	933 E. 56th St.	Chicago	IL	60637	

## 6	NA 111 Nowelo St.	Hilo	HI	96720
## 7	NA 2145 Sheridan Rd	Evanston	IL	60208-3112
## 8	NA 933 E. 56th St.	Chicago	IL	60637
## 9	NA 174 W. 18th Ave.	Columbus	OH	43210
## 10	NA 5000 Forbes Ave.	Pittsburgh	PA	15213
## 11	NA 933 E. 56th St.	Chicago	IL	60637
## 12	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 13	NA 725 W. Commonwealth Ave.	Boston	MA	02215
## 14	NA 725 W. Commonwealth Ave.	Boston	MA	02215
## 15	NA 725 W. Commonwealth Ave.	Boston	MA	02215
## 16	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 17	NA 5640 S. Ellis Ave.	Chicago	IL	60637
## 18	NA 5640 S. Ellis Ave.	Chicago	IL	60637
## 19	NA 1002 W. Green St.	Urbana	IL	61801
## 20	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 21	NA 4201 Wilson Blvd	Arlington	VA	22230
## 22	NA 174 W. 18th Ave.	Columbus	OH	43210
## 23	NA 933 E. 56th St.	Chicago	IL	60637
## 24	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 25	NA 5000 Forbes Ave.	Pittsburgh	PA	15213
## 26	NA 2145 Sheridan Rd	Evanston	IL	60208-3112
## 27	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 28	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 29	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 30	NA 5000 Forbes Ave.	Pittsburgh	PA	15213
## 31	NA 933 E. 56th St.	Chicago	IL	60637
## 32	NA 5640 S. Ellis Ave.	Chicago	IL	60637
## 33	NA 2145 Sheridan Rd	Evanston	IL	60208-3112
## 34	NA 8730 W. Mountain View Ln	Littleton	CO	80125
## 35	NA 933 E. 56th St.	Chicago	IL	60637
## 36	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 37	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 38	NA 933 N. Cherry St.	Tucson	AZ	85721
## 39	NA 5000 Forbes Ave.	Pittsburgh	PA	15213
## 40	NA 373 W. Geneva St.	Wms. Bay	WI	53191
## 41	NA 791 Holmdel-Keyport Rd.	Holmdel	NY	07733-1988
## 42	NA 5640 S. Ellis Ave.	Chicago	IL	60637

Problem 4:

```
url <- "https://github.com/fonnesbeck/Bios6301/raw/master/datasets/haart.csv"
haart_df <- read.csv(url)[,c('death', 'weight', 'hemoglobin', 'cd4baseline')]
coef(summary(glm(death ~ ., data=haart_df, family=binomial(logit))))
```

##		Estimate	Std. Error	z value	Pr(> z)
##	(Intercept)	3.576411744	1.226870535	2.915069	0.0035561039
##	weight	-0.046210552	0.022556001	-2.048703	0.0404911395
##	hemoglobin	-0.350642786	0.105064078	-3.337418	0.0008456055
##	cd4baseline	0.002092582	0.001811959	1.154872	0.2481427160

```
myfun <- function(dat, response) {
  form <- as.formula(response ~ .)
  coef(summary(glm(form, data=dat, family=binomial(logit))))
}
```

```
}
myfun(haart_df, haart_df$death)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
##               Estimate Std. Error      z value Pr(>|z|)
## (Intercept) -2.656607e+01 115935.1724 -2.291459e-04 0.9998172
## death       5.313214e+01  69028.4188  7.697140e-04 0.9993859
## weight      -2.439288e-15   1939.0571 -1.257976e-18 1.0000000
## hemoglobin   2.781698e-14   9774.8190  2.845780e-18 1.0000000
## cd4baseline  2.035788e-16    184.0846  1.105898e-18 1.0000000
```

The problem is that death is being treated as a predictor, when it is supposed to be the outcome.

```
myfun_fix <- function(dat, response) {
  response<- deparse(substitute(response))
  form <- as.formula(paste(response, "~."))
  coef(summary(glm(form, data=dat, family=binomial(logit))))
}
myfun_fix(haart_df, death)
```

```
##               Estimate Std. Error      z value      Pr(>|z|)
## (Intercept)  3.576411744 1.226870535  2.915069 0.0035561039
## weight      -0.046210552 0.022556001 -2.048703 0.0404911395
## hemoglobin  -0.350642786 0.105064078 -3.337418 0.0008456055
## cd4baseline  0.002092582 0.001811959  1.154872 0.2481427160
```